Glen, A. (2007) *Order and quasiperiodicity in episturmian words.* In: Sixth International Conference on Words, 17 - 21 September, Marseille, France.

Order and quasiperiodicity in episturmian words

Amy Glen*

LaCIM, Université du Québec à Montréal,
C.P. 8888, succursale Centre-ville, Montréal, Québec, CANADA, H3C 3P8

Extended Abstract†

July 13, 2007

1 Introduction

In this paper, we build upon previous work concerning inequalities characterizing Sturmian and episturmian words – see [19, 29, 30, 14, 16]. First let us recall from [29] the following notion relating to *lexicographic order*. Let $\mathcal{A}$ be a *totally ordered* finite alphabet consisting of at least two letters. To any infinite word $\mathbf{x}$ over $\mathcal{A}$, we can associate two infinite words $\min(\mathbf{x})$ and $\max(\mathbf{x})$ such that any prefix of $\min(\mathbf{x})$ (resp. $\max(\mathbf{x})$) is the *lexicographically* smallest (resp. greatest) amongst the factors of $\mathbf{x}$ of the same length. More precisely, if we denote by $\min(\mathbf{x}|k)$ (resp. $\max(\mathbf{x}|k)$) the lexicographically smallest (resp. greatest) factor of $\mathbf{x}$ of length k for the given order, then $\min(\mathbf{x}|k)$ and $\max(\mathbf{x}|k)$ are clearly prefixes of the respective words $\min(\mathbf{x}|k+1)$ and $\max(\mathbf{x}|k+1)$. So we can define, by taking limits, the following two infinite words

$$\min(\mathbf{x}) = \lim_{k\to\infty} \min(\mathbf{x}|k) \quad \text{and} \quad \max(\mathbf{x}) = \lim_{k\to\infty} \max(\mathbf{x}|k).$$

An important point here is that, for any aperiodic infinite word $\mathbf{x}$ which is uniformly recurrent, $\min(\mathbf{x})$ is an *infinite Lyndon word*, i.e., it is (strictly) lexicographically smaller than all of its proper suffixes for the given order on $\mathcal{A}$.

In 2003, Pirillo [28] (also see [29]) proved that, for infinite words $\mathbf{s}$ on a 2-letter alphabet $\{a, b\}$ with $a < b$, the inequality $a\mathbf{s} \le \min(\mathbf{s}) \le \max(\mathbf{s}) \le b\mathbf{s}$ characterizes *standard Sturmian words* (aperiodic and periodic). Equivalently, an infinite sequence $\mathbf{s} = (s_n)_{n\ge0}$ over $\{a, b\}$ is standard Sturmian if and only if

$$a\mathbf{s} \le \mathrm{T}^k(\mathbf{s}) \le b\mathbf{s}, \quad \text{for all } k \ge 0, \tag{1.1}$$

where T^k is the k-th iterate of the *shift map*: $\mathrm{T}^k((s_n)_{n\ge0}) = (s_{n+k})_{n\ge0}$ (*cf.* our analogue for episturmian sequences – Corollary 3.2). Actually, this result was known much earlier, dating back to the work of P. Veerman [38, 39] in the mid 80's. Since that time, these 'Sturmian inequalities' have been rediscovered numerous times under different guises, as discussed in our survey paper [4].

In the case of an arbitrary finite alphabet $\mathcal{A}$, Pirillo [29] generalized the above inequalities by proving that an infinite word $\mathbf{s}$ over $\mathcal{A}$ is *standard episturmian* (or *epistandard* for short) if and only if, for any lexicographic order, we have

$$a\mathbf{s} \le \min(\mathbf{s}) \quad \text{where } a = \min(\mathcal{A}). \tag{1.2}$$

Moreover, $\mathbf{s}$ is a *strict* epistandard word (i.e., a *standard Arnoux-Rauzy sequence* [7, 37]) if and only if (1.2) holds with strict equality for any order [19]. In a similar spirit, Glen, Justin, and Pirillo [16] recently proved the following characterization of all *episturmian words*.

Proposition 1.1. [16] *A recurrent infinite word* $\mathbf{t}$ *over* $\mathcal{A}$ *is episturmian if and only if there exists an infinite word* $\mathbf{s}$ *such that, for any lexicographic order, we have* $a\mathbf{s} \le \min(\mathbf{t})$ *where* $a = \min(\mathcal{A})$. □

From the proof of the above result, it is not immediately clear what form is taken by the infinite word $\mathbf{s}$, if it exists. We now prove further (in Section 3) that $\mathbf{s}$ is in fact the 'unique' epistandard word with the same set of factors as $\mathbf{t}$, i.e., the corresponding epistandard word in the *shift orbit closure* of $\mathbf{t}$ (see Section 2).

As the title of this paper suggests, these results have a connection with *quasiperiodicity*. What exactly? Well, roughly speaking, an infinite word $\mathbf{x}$ is *quasiperiodic* if there exists a finite word u such that the

*E-mail: amy.glen@gmail.com (with the support of CRM-ISM-LaCIM)

†Same as the full version, but all proofs (and some lemmas) have been omitted.

occurrences of u in $\mathbf{x}$ entirely cover $\mathbf{x}$, i.e., every position of $\mathbf{x}$ falls within some occurrence of u in $\mathbf{x}$ (e.g., see Marcus [26]).

At first glance, it would seem that most, if not all, (epi)sturmian words are quasiperiodic. Certainly, one would be quick to draw this conclusion if only epistandard words were considered, since all such words are quasiperiodic (see Theorem 4.5). But, alas, Levé and Richomme [22] have shown via an explicit example that not every Sturmian word is quasiperiodic. Of course, a tempting question to ask is then: "Which (epi)sturmian words are not quasiperiodic?" Recently, in [23], the same two authors answered this question (which was first posed by Marcus [26]) for Sturmian words, by characterizing those that are not quasiperiodic. Specifically, they proved that an (aperiodic) Sturmian word is not quasiperiodic if and only if it can be decomposed infinitely over the set of morphisms $\{\psi_a, \bar{\psi}_b\}$ or $\{\psi_b, \bar{\psi}_a\}$ (defined in Section 2.1). Furthermore, a Sturmian word is not quasiperiodic if and only if it is an infinite Lyndon word for the given order on $\{a, b\}$ ($a < b$ or $b < a$). Of particular interest here is the following consequence: a Sturmian word is an infinite Lyndon word if and only if it can be infinitely decomposed over $\{\psi_a, \bar{\psi}_b\}$ for $a < b$, or $\{\psi_b, \bar{\psi}_a\}$ for $b < a$.

Here we prove that an $\mathcal{A}$-*strict* episturmian word $\mathbf{t}$ is an infinite Lyndon word if and only if $\mathbf{t} = a\mathbf{s}$ where $a = \min(\mathcal{A})$ for the given order on $\mathcal{A}$ and $\mathbf{s}$ is an (aperiodic) $\mathcal{A}$-strict epistandard word (see Theorem 3.6). This theorem follows quite simply from our refinement of Proposition 1.1 in Section 3. It easily follows, too, that an $\mathcal{A}$-strict episturmian word is an infinite Lyndon word if and only if it is infinitely decomposable over $\{\psi_a, \bar{\psi}_x \mid x \in \mathcal{A} \setminus \{a\}\}$, where $a = \min(\mathcal{A})$ for the given order on $\mathcal{A}$ (see Theorem 3.8). In light of these latter results and those of Levé and Richomme [23] for Sturmian words, it is natural to conjecture that the non-quasiperiodic episturmian words are precisely those that are infinite Lyndon words. This assertion is not correct; in fact, it is rather far from the truth, as we show in Section 4. Specifically, we prove that an episturmian word is not quasiperiodic if it is directed by a *regular wavy word* (see Theorem 4.21). This shows that there is a much wider class of episturmian words that are not quasiperiodic, besides those that are infinite Lyndon words. From these results and others in Section 4, we establish a characterization of the (non)-quasiperiodic episturmian words in terms of their directive words (Theorems 4.24–4.25). Whereas *Sturmian morphisms* were the main tool used in [23], we take a different approach by using the notion of *return words* to obtain an equivalent definition of quasiperiodicity (see Theorem 4.2). This allows us to completely describe all of the *quasiperiods* of an epistandard word. Directive words also play a key role in our study of quasiperiodicity.

Lastly, in Section 5, we use our main result on episturmian Lyndon words (Theorem 3.6), together with a result of W. Parry [27], to prove a simple characterization of so-called *self-episturmian numbers*.

Note. In these present proceedings of the Words '07 conference, F. Levé and G. Richomme establish another characterization of the quasiperiodic episturmian words using morphic decompositions. They also characterize the *epistandard morphisms* that are *strongly quasiperiodic*.

1.1 Notation & terminology

Let $\mathcal{A}$ be an arbitrary finite alphabet. For any finite or infinite word w over $\mathcal{A}$, $F(w)$ denotes the set of all its factors, and $F_n(w)$ denotes the set of all factors of w of length $n \in \mathbb{N}$ (where $|w| \geq n$ for w finite). Moreover, the *alphabet* of w is $\mathrm{Alph}(w) := F(w) \cap \mathcal{A}$ and, if w is infinite, we denote by $\mathrm{Ult}(w)$ the set of all letters occurring infinitely often in w. Any two infinite words $\mathbf{x}$, $\mathbf{y} \in \mathcal{A}^\omega$ are said to be *equivalent* if $F(\mathbf{x}) = F(\mathbf{y})$. The empty word is denoted by ε. For other basic notions and concepts in combinatorics on words, the reader may wish to consult Lothaire [24, 25] for example.

2 Facts about episturmian words

An interesting natural generalization of the aperiodic *Sturmian words* (e.g., see [25, Chapter 2]) to a finite alphabet is the family of *Arnoux-Rauzy sequences*, the study of which began in [7] (also see [19, 37] for example). More recently, a slightly wider class of infinite words, aptly called *episturmian words*, was introduced by Droubay, Justin, and Pirillo [11] (also see [13, 18, 20, 21] for instance). In this section, we recall some basic definitions and properties relating to episturmian words which are needed later in the paper. For the most part, we follow the notation and terminology of [11, 18, 20, 16].

Definition 2.1. [18] An infinite word $\mathbf{t} \in \mathcal{A}^\omega$ is ***episturmian*** if $F(\mathbf{t})$ is closed under reversal and $\mathbf{t}$ has at most one right (or equivalently left) special factor of each length. Moreover, an episturmian word is ***standard*** if all of its left special factors are prefixes of it.

This definition gives the aperiodic, as well as periodic, Sturmian words when $|\mathcal{A}| = 2$.

Note. Hereafter, we refer to a standard episturmian word as an ***epistandard word***, for simplicity.

Epistandard words were characterized in [11, 18] using the concept of the *palindromic right-closure* $w^{(+)}$ of a finite word w, which is the (unique) shortest palindrome having w as a prefix (see [10]). That is, $w^{(+)} = wv^{-1}\widetilde{w}$ where v is the longest palindromic suffix of w.

Proposition 2.2. [11] *An infinite word* $\mathbf{s} \in \mathcal{A}^\omega$ *is epistandard if and only if there exists an infinite word* $\Delta = x_1x_2x_3\ldots$ *(*$x_i \in \mathcal{A}$*), called the* directive word *of* $\mathbf{s}$*, such that* $\mathbf{s} = \lim_{n\to\infty} u_n$*, with*

$$u_1 = \varepsilon, \quad u_{i+1} = (u_ix_i)^{(+)}, \quad \textit{for all } i \geq 1. \tag{2.1}$$

Note. Δ *uniquely* determines the epistandard word $\mathbf{s}$.

This characterization extends to the case of an arbitrary finite alphabet a construction given in [10] for all *standard Sturmian words*. An important point is that an epistandard word can be constructed as a limit of an infinite sequence of palindromes $(u_i)_{i\geq 0}$, which are exactly its palindromic prefixes (by construction).

2.1 Episturmian morphisms

Let $a \in \mathcal{A}$ and denote by ψ_a, $\bar{\psi}_a$ the morphisms on $\mathcal{A}$ defined by

$$\psi_a : \begin{cases} a & \mapsto & a \\ x & \mapsto & ax \end{cases}, \quad \bar{\psi}_a : \begin{cases} a & \mapsto & a \\ x & \mapsto & xa \end{cases} \quad \text{for all } x \in \mathcal{A} \setminus \{a\}.$$

Together with the permutations of the alphabet, all of the morphisms ψ_a, $\bar{\psi}_a$ generate by composition the monoid of *episturmian morphisms* [11, 18]. The monoid of *epistandard morphisms* [34] is generated by all the ψ_a and permutations on $\mathcal{A}$. Further, the submonoid of *pure episturmian morphisms* (resp. *pure epistandard morphisms*) is generated by the ψ_a and $\bar{\psi}_a$ only (resp. ψ_a only).

2.1.1 Relation with episturmian words

Let $\bar{\mathcal{A}} = \{\bar{x} \mid x \in \mathcal{A}\}$. A letter $\bar{x}$ is considered to be x with *spin* 1, whilst x itself has spin 0. An infinite *spinned word* is an element of $(\mathcal{A} \cup \bar{\mathcal{A}})^\omega$. When the spins are not explicitly given, we write such an infinite word in the form $\breve{x}_1\breve{x}_2\breve{x}_3\cdots$, where $\breve{x}_i = x_i$ if x_i has spin 0, or $\breve{x}_i = \bar{x}_i$ if x_i has spin 1.

In terms of episturmian morphisms, Justin and Prillo [18] proved the following insightful characterization of episturmian words, which shows that any episturmian word can be *infinitely decomposed* over the pure episturmian morphisms ψ_x, $\bar{\psi}_x$ $(x \in \mathcal{A})$.

Proposition 2.3. [18] *An infinite word* $\mathbf{t} \in \mathcal{A}^\omega$ *is episturmian if and only if there exists an infinite spinned directive word* $\breve{\Delta} = \breve{x}_1\breve{x}_2\cdots$ *(*$x_i \in \mathcal{A}$*) and an infinite sequence* $(\mathbf{t}^{(i)})_{i\geq 0}$ *of recurrent infinite words such that*

$$\mathbf{t}^{(0)} = \mathbf{t} \quad \textit{and} \quad \mathbf{t}^{(i-1)} = \psi_{x_i}(\mathbf{t}^{(i)}) \quad \textit{or} \quad \mathbf{t}^{(i-1)} = \bar{\psi}_{x_i}(\mathbf{t}^{(i)}), \quad \textit{for all } i > 0,$$

according to the spin 0 *or* 1 *of* $\breve{x}_i$*, respectively. Moreover, each* $\mathbf{t}^{(i)}$ *is an episturmian word directed by* $\mathrm{T}^i(\breve{\Delta}) = \breve{x}_{i+1}\breve{x}_{i+2}\cdots$ *and is equivalent to the (unique) epistandard word* $\mathbf{s}^{(i)}$ *directed by* $x_{i+1}x_{i+2}\cdots$. □

Remark 2.4. For any episturmian word $\mathbf{t}$ and spinned infinite word $\breve{\Delta}$ satisfying the conditions of the above theorem, we say that $\breve{\Delta}$ is a *spinned directive word* for $\mathbf{t}$ or $\mathbf{t}$ is *directed by* $\breve{\Delta}$. In general, a spinned word directing an episturmian word is not unique (see [18, 20] or §4.3.1). For example, the *Tribonacci word* is directed by $(abc)^\omega$ and also $(abc)^n\bar{a}\bar{b}\bar{c}(a\bar{b}\bar{c})^\omega$ for each $n \geq 0$, as well as infinitely many other spinned words.

Notation. From now on, unless stated otherwise, the notation $\Delta = x_1x_2x_3\cdots$ $(x_i \in \mathcal{A})$ will remain for epistandard $\mathbf{s}$, and any equivalent episturmian word $\mathbf{t}$ with spinned directive word $\breve{\Delta} = \breve{x}_1\breve{x}_2\breve{x}_3\cdots$.

2.1.2 Notation for pure episturmian morphisms

To a finite word $w = x_1x_2\cdots x_n$ $(x_i \in \mathcal{A})$, we associate the pure epistandard morphism $\mu_w := \psi_{x_1}\psi_{x_2}\cdots\psi_{x_n}$. When considering a spinned version $\breve{w} = \breve{x}_1\breve{x}_2\cdots\breve{x}_n$ of w, the corresponding pure episturmian morphism $\mu_{\breve{w}}$ is obtained from μ_w by replacing ψ_{x_i} by $\bar{\psi}_{x_i}$ when $\breve{x}_i = \bar{x}_i$ (spin 1). In particular, $\mu_x = \psi_x$ and $\mu_{\bar{x}} = \bar{\psi}_x$. We write $\mu_n = \mu_w$ and $\breve{\mu}_n = \mu_{\breve{w}}$, and define $\mu_0 = \breve{\mu}_0 = \mathrm{Id}$. Viewing $w = x_1x_2\cdots x_n$ as a prefix of the directive word Δ, it is clear from Proposition 2.3 that the words $h_{n-1} := \mu_{n-1}(x_n)$, $n \geq 1$, are prefixes of the epistandard word $\mathbf{s}$. For the palindromic prefixes $(u_i)_{i\geq 1}$ given by (2.1), we have the following useful formula [18]

$$u_{i+1} = h_{i-1}u_i \quad \text{for } i > 0. \tag{2.2}$$

2.1.3 Strictness

An epistandard word $\mathbf{s} \in \mathcal{A}^\omega$, or any equivalent (episturmian) word $\mathbf{t}$, is said to be $\mathcal{B}$-*strict* (or k-*strict* if $|\mathcal{B}| = k$, or *strict* if $\mathcal{B}$ is understood) if $\mathrm{Alph}(\Delta) = \mathrm{Ult}(\Delta) = \mathcal{B} \subseteq \mathcal{A}$. That is, an episturmian word is strict if every letter in its alphabet occurs infinitely often in its directive word. The k-strict episturmian words are exactly the k-letter Arnoux-Rauzy sequences. In particular, the 2-strict episturmian words correspond to the (aperiodic) Sturmian words.

Note. An episturmian word is periodic if and only if $|\mathrm{Ult}(\Delta)| = 1$ (see [18, Proposition 2.9]). More precisely, a periodic episturmian word takes the form $(\mu_{\breve{w}}(x))^\omega$ for some finite spinned word $\breve{w}$ and letter x.

2.2 Episturmian orbits

The *shift orbit* of an infinite word $\mathbf{x} \in \mathcal{A}^\omega$ is the set $\mathcal{O}(\mathbf{x}) = \{\mathrm{T}^i(\mathbf{x}) \mid i \geq 0\}$ and its *closure* is given by $\overline{\mathcal{O}}(\mathbf{x}) = \left\{\mathbf{y} \in \mathcal{A}^\omega \mid \mathrm{Pref}(\mathbf{y}) \subseteq \bigcup_{i\geq 0} \mathrm{Pref}(\mathrm{T}^i(\mathbf{x}))\right\}$, where $\mathrm{Pref}(w)$ denotes the set of prefixes of a finite or infinite word w. Here, we use the phrase *episturmian orbit* to refer to the shift orbit closure $\overline{\mathcal{O}}(\mathbf{t})$ of an episturmian word $\mathbf{t}$ (*cf.* [37]). Any infinite word $\mathbf{x}$ in $\overline{\mathcal{O}}(\mathbf{t})$ is *equivalent* to $\mathbf{t}$ in the sense that $\mathbf{x}$ has the same set of factors as $\mathbf{t}$. Certainly, for each n, we have $F_n(\overline{\mathcal{O}}(\mathbf{t})) = F_n(\mathbf{t})$ since the uniform recurrence property [11] of episturmian words implies that $F_n(\mathbf{x}) = F_n(\mathbf{t})$ for each $\mathbf{x} \in \overline{\mathcal{O}}(\mathbf{t})$. In other words, an episturmian orbit $\overline{\mathcal{O}}(\mathbf{t})$ is a *minimal dynamical system* (e.g., see [31]) as it contains all of the episturmian words with the same set of factors.

By Proposition 2.3, any episturmian word $\mathbf{t}$ directed by a spinned word of the form $\breve{\Delta} = \breve{x}_1\breve{x}_2\breve{x}_3\cdots$ is equivalent to the unique epistandard word $\mathbf{s}$ directed by $\Delta = x_1x_2x_3\cdots$. In particular, every *aperiodic* episturmian orbit $\overline{\mathcal{O}}(\mathbf{t})$ contains a unique epistandard word $\mathbf{s}$, which can be viewed as the (unique) accumulation point of the set of all left special factors of $\mathbf{t}$ (by Proposition 2.2). As the directive word Δ completely determines the aperiodic epistandard word $\mathbf{s}$, and hence the set of factors $F(\overline{\mathcal{O}}(\mathbf{s})) = F(\overline{\mathcal{O}}(\mathbf{t}))$, any two aperiodic episturmian orbits $\overline{\mathcal{O}}(\mathbf{t})$, $\overline{\mathcal{O}}(\mathbf{t}')$, with respective directive words $\breve{\Delta}_{\mathbf{t}}$, $\breve{\Delta}_{\mathbf{t}'}$, are equal if and only if $\Delta_{\mathbf{t}} = \Delta_{\mathbf{t}'}$ (all spins 0). On the other hand, a *periodic* episturmian orbit does not contain a unique epistandard word, but rather two epistandard words, $\mathbf{s}_1$ and $\mathbf{s}_2$ say. For example, consider $\mathbf{s}_1 = (bcba)^\omega$ directed by $\Delta_1 = bca^\omega$ and $\mathbf{s}_2 = (babc)^\omega$ directed by $\Delta_2 = bac^\omega$. Both are epistandard words with the same set of factors, and hence they are in the same episturmian orbit.

Remark 2.5. Given any episturmian word $\mathbf{t}$ directed by $\breve{\Delta}$, we can associate to its (aperiodic or periodic) episturmian orbit $\overline{\mathcal{O}}(\mathbf{t})$ a unique epistandard word $\mathbf{s}$, which is the one directed by Δ. We often use this fact implicitly in what follows.

We prove the following important theorem which is useful for establishing our main results in Sections 3–4.

Theorem 2.6. *Suppose* $\mathbf{s}$ *is an epistandard word directed by* Δ *and let* a *be a letter. Then* $a\mathbf{s}$ *is an episturmian word if and only if* $a \in \mathrm{Ult}(\Delta)$. *Moreover,* $a\mathbf{s}$ *is the (unique) episturmian word in* $\overline{\mathcal{O}}(\mathbf{s})$ *directed by* $\bar{\Delta}_a$ *which is* Δ *with all spins* 1 *except when* $x_i = a$. □

Notation. Hereafter, we continue to denote by $\bar{\Delta}_a$ the directive word of $a\mathbf{s}$, with the assumption that $a \in \mathrm{Ult}(\Delta)$ (otherwise $a\mathbf{s}$ is not episturmian, by Theorem 2.6).

3 Extremal properties of episturmian words

3.1 A refinement of Proposition 1.1

Theorem 3.1. *For any recurrent infinite word* $\mathbf{t} \in \mathcal{A}^\omega$*, the following properties are equivalent.*

i) $\mathbf{t}$ *is an episturmian word directed by* $\breve{\Delta}$.

ii) There exists an infinite word $\mathbf{x}$ *such that, for any lexicographic order, we have* $a\mathbf{x} \leq \min(\mathbf{t})$ *where* $a = \min(\mathcal{A})$.

Moreover, $\mathbf{x}$ *is the (unique) epistandard word in* $\overline{\mathcal{O}}(\mathbf{t})$ *directed by* Δ*, with the property that* $a\mathbf{x} = \min(\mathbf{t})$ *if and only if* $a = \min(\mathcal{A})$ *belongs to* $\mathrm{Ult}(\Delta)$. □

From Theorem 3.1, we quite easily deduce the following corollary, which is an analogue of the Sturmian inequalities (1.1) given in the introduction. We state only the fact that 'episturmianicity' implies the inequalities, as it is sufficient for our purposes. Obviously the converse holds too (by Theorem 3.1).

Corollary 3.2. *Suppose* $\mathbf{t} \in \mathcal{A}^\omega$ *is an episturmian word directed by* $\breve{\Delta}$ *and* $\mathbf{s}$ *is the unique epistandard word in* $\overline{\mathcal{O}}(\mathbf{t})$ *with directive word* Δ*. Then, for any lexicographic order,* $a\mathbf{s} \leq \mathrm{T}^k(\mathbf{t})$ *(resp.* $\mathrm{T}^k(\mathbf{t}) \leq b\mathbf{s}$*) for all* $k \geq 0$*, where* $a = \min(\mathcal{A})$ *(resp.* $b = \max(\mathcal{A})$*). Moreover,* $\mathbf{s}$ *has the property that* $a\mathbf{s} = \inf_k \mathrm{T}^k(\mathbf{t})$ *(resp.* $b\mathbf{s} = \sup_k \mathrm{T}^k(\mathbf{t})$*) if and only if* $a = \min(\mathcal{A})$ *(resp.* $b = \max(\mathcal{A})$*) belongs to* $\mathrm{Ult}(\Delta)$*.* □

Remark 3.3. A noteworthy fact is that an episturmian orbit $\overline{\mathcal{O}}(\mathbf{t})$ is 'dominated' by $b\mathbf{s}$ for any order such that $b = \max(\mathcal{A})$ belongs to $\mathrm{Ult}(\Delta)$.

3.2 Episturmian Lyndon words

From the preceding results, we easily obtain the following characterization of the strict episturmian words that are infinite Lyndon words (see Theorem 3.6 below). We also characterize the strict episturmian Lyndon words via morphisms.

Let us first recall that a non-empty finite word w over $\mathcal{A}$ is a *Lyndon word* if it is lexicographically smaller than all of its proper suffixes for the given order $<$ on $\mathcal{A}$. Equivalently, w is the lexicographically smallest word in its conjugacy class; that is, $w < vu$ for all non-empty words u, v such that $w = uv$. The first of these definitions extends to infinite words: an infinite word over $\mathcal{A}$ is an *infinite Lyndon word* if and only if it is (strictly) lexicographically smaller than all of its proper suffixes for the given order on $\mathcal{A}$. Note that any periodic infinite word, say $\mathbf{x} = v^\omega$ ($v \in \mathcal{A}^+$), cannot be an infinite Lyndon word. Indeed, even if v is itself a finite Lyndon word, $\mathbf{x}$ is lexicographically smaller than *or equal to* all of its proper suffixes, but not strictly smaller, since $\mathrm{T}^{m|v|}(\mathbf{x}) = \mathbf{x}$ for all $m \geq 1$.

Note. In this section, we assume that $|\mathcal{A}| \geq 2$ since on a 1-letter alphabet there are no infinite Lyndon words.

Remark 3.4. From now on, when we refer to an *aperiodic* episturmian word $\mathbf{t}$, it is important to remember that this means $|\mathrm{Ult}(\Delta)| \geq 2$.

Proposition 3.5. *Suppose* $\mathbf{t} = a\mathbf{s}$ *where* $a = \min(\mathcal{A})$ *for the given order on* $\mathcal{A} = \mathrm{Alph}(\mathbf{t})$ *and* $\mathbf{s}$ *is an aperiodic epistandard word with* $a \in \mathrm{Ult}(\Delta)$*. Then* $\mathbf{t}$ *is an episturmian Lyndon word.* □

Theorem 3.6. *An* $\mathcal{A}$*-strict episturmian word* $\mathbf{t}$ *is an infinite Lyndon word if and only if* $\mathbf{t} = a\mathbf{s}$ *where* $a = \min(\mathcal{A})$ *for the given order on* $\mathcal{A}$ *and* $\mathbf{s}$ *is an (aperiodic)* $\mathcal{A}$*-strict epistandard word.* □

Given an $\mathcal{A}$-strict epistandard word $\mathbf{s}$, Theorem 3.6 shows that there exist exactly $|\mathcal{A}| \geq 2$ infinite Lyndon words that are equivalent to $\mathbf{s}$. That is, for any order with $\min(\mathcal{A}) = a$, $\overline{\mathcal{O}}(\mathbf{s})$ contains a unique infinite Lyndon word beginning with a, namely $a\mathbf{s}$. By Theorem 2.6, $a\mathbf{s}$ is the episturmian word directed by $\bar{\Delta}_a$ which is Δ with all spins 1, except when $x_i = a$. In particular, the following result is readily deduced from Theorems 2.6 and 3.6.

Corollary 3.7. *An* $\mathcal{A}$*-strict episturmian word* $\mathbf{t}$ *is an infinite Lyndon word if and only if* $\mathbf{t}$ *is directed by* $\bar{\Delta}_a$ *where* $a = \min(\mathcal{A})$ *for the given order on* $\mathcal{A}$*.* □

Theorem 3.8. *An* $\mathcal{A}$*-strict episturmian word* $\mathbf{t}$ *is an infinite Lyndon word if and only if it can be infinitely decomposed over* $\{\psi_a, \bar{\psi}_x \mid x \in \mathcal{A} \setminus \{a\}\}$ *where* $a = \min(\mathcal{A})$ *for the given order on* $\mathcal{A}$*.* □

Note. Episturmian morphisms that preserve finite and infinite Lyndon words were characterized in [35, 36].

4 Quasiperiodicity

We take from [23] the definitions of finite and infinite quasiperiodic words, as follows (also see [26]). A finite word u *covers* another finite word $w \neq u$ if for every $i \in \{1, \ldots, |w|\}$, there exists $j \in \{1, \ldots, |u|\}$ such that there is an occurrence of u starting a position $i - j + 1$ in the word w. We say that u is a *quasiperiod* of w, and that w is u*-quasiperiodic*, or simply *quasiperiodic*. That is, a finite word w is *quasiperiodic* if there exists a word $u \neq w$ such that the occurrences of u in w entirely cover w, i.e., every position of w falls within some occurrence of u in w. For example, the finite word $ababa$ is aba-quasiperiodic. (See [6] and also [5] for a brief survey of quasiperiodicity in 'strings'.) Similarly, an infinite word $\mathbf{x}$ is *quasiperiodic* if there exists a finite word u such that the occurrences of u in $\mathbf{x}$ entirely cover $\mathbf{x}$. Or, more precisely, $\mathbf{x}$ is *quasiperiodic* if there exist a finite word u and words $(p_n)_{n\geq 0}$ such that $p_0 = \varepsilon$, $|p_n| < |p_{n+1}| - |p_n| \leq |u|$, and $p_n u$ is a prefix of $\mathbf{x}$ for all $n \geq 0$. As for finite words, we say that u *covers* $\mathbf{x}$ and that u is a *quasiperiod* of $\mathbf{x}$. Necessarily, any quasiperiod of a quasiperiodic word must be a prefix of it.

Recently, Levé and Richomme [23] characterized the non-quasperiodic Sturmian words. The aim of this section is to do the same for episturmian words. Our methods here are different to those utilized in [23]. Instead of using (epi)sturmian morphisms as the main tool, we approach the problem by using the notion of return words to give an equivalent definition of quasiperiodicity.

4.1 Return words & quasiperiodicity

We now recall the notion of a *return word*, which was introduced independently by Durand [12] and Holton-Zamboni [17] when studying primitive substitutive sequences.

Definition 4.1. Let v be a recurrent factor of $\mathbf{y} \in \mathcal{A}^\omega$, starting at positions $n_1 < n_2 < n_3 \cdots$. Then each word $r_i = y_{n_i} y_{n_i+1} \cdots y_{n_{i+1}-1}$ is called a ***return word*** to v in $\mathbf{y}$.

That is, we define the set $\mathcal{R}_v(\mathbf{y})$ of return words to v to be the set of all distinct words beginning at an occurrence of v and ending exactly before the next occurrence of v in $\mathbf{y}$. Thus, a return word to v in $\mathbf{y}$ is a non-empty factor r such that v is a prefix of rv and rv contains exactly two occurrences of v. We call rv a *complete return word* of v [21]. Clearly, $\mathcal{R}_v(\mathbf{y})$ is finite for all $v \in F(\mathbf{y})$ if and only if $\mathbf{y}$ is uniformly recurrent. This is true for episturmian words since they are uniformly recurrent [11].

Note that a return word to v in $\mathbf{y}$ always has v as a prefix or is a prefix of v. In particular, observe that a return word to v is not necessarily longer than v, in which case v has *overlapping* occurrences in $\mathbf{y}$ (i.e., $vz^{-1}v \in F(\mathbf{y})$ is a complete return word of v for some non-empty word z). We say that v has *adjacent* occurrences in $\mathbf{y}$ if vv is a factor of $\mathbf{y}$. In this case, if v is *primitive*, then v is a return word to itself; otherwise, the corresponding return word is the *primitive root* of v.

We prove the following equivalent definition of a quasiperiodic infinite word in terms of return words.

Theorem 4.2. *An infinite word* $\mathbf{x}$ *is quasiperiodic if and only if there exists a recurrent prefix* w *of* $\mathbf{x}$ *such that all of the return words to* w *in* $\mathbf{x}$ *have length at most* $|w|$*, in which case* w *is a quasiperiod of* $\mathbf{x}$*. Moreover, the shortest such prefix* w *is the smallest quasiperiod of* $\mathbf{x}$*.* □

Remark 4.3. A noteworthy fact is that a quasiperiodic infinite word is not necessarily recurrent [26], although it must have a prefix that is recurrent in it.

4.2 Quasiperiodic episturmian words

4.2.1 Quasiperiods

In this section, we extend Lemma 6.3 in [22] by showing that any epistandard word $\mathbf{s}$ is quasiperiodic. Even further, we completely describe all of the quasiperiods. The following proposition is useful.

Proposition 4.4 (Justin-Vuillon [21])**.** *Let* $\mathbf{s}$ *be an epistandard word and* $v \in F(\mathbf{s})$*. If* u_{n+1} *is the shortest palindromic prefix of* $\mathbf{s}$ *containing* v *with* $u_{n+1} = fvg$*, then the return words to* v *are given by* $f^{-1}\mu_n(x)f$ *where* $x \in \mathrm{Alph}(x_{n+1}x_{n+2}\cdots)$*. Moreover, the corresponding complete return words of* v *are the words* $f^{-1}(u_{n+1}x)^{(+)}g^{-1}$*.* □

As in [21, 18], let us define $P(i) = \sup\{j < i \mid x_j = x_i\}$ if this number exists, undefined otherwise. Then, by the definitions of palindromic closure and the palindromes $(u_i)_{i\geq 1}$, it follows that $u_{n+1} = u_n x_n u_n$ (whence $\mu_{n-1}(x_n) = u_n x_n$) if x_n does not occur in u_n, and $u_{n+1} = u_n u_{P(n)}^{-1} u_n$ (whence $\mu_{n-1}(x_n) = u_n u_{P(n)}^{-1}$) if x_n occurs in u_n. Therefore, the length of the longest return word r_{n+1} to u_{n+1} in $\mathbf{s}$ is given by

$$|r_{n+1}| = \begin{cases} |u_{n+1}| + 1 & \text{if some } x \in \mathrm{Alph}(\mathbf{s}) \text{ does not occur in } u_{n+1}, \\ |u_{n+1}| - |u_p| & \text{otherwise,} \end{cases}$$

where $p = \inf\{P(i) \mid i \geq n+1\}$ (also see [21, Lemma 5.6]).

Now, let m be minimal such that $\mathrm{Alph}(x_1x_2\cdots x_m) = \mathrm{Alph}(\mathbf{s})$. Then $u_{m+1} = u_m x_m u_m$ is the shortest palindromic prefix of $\mathbf{s}$ such that $\mathrm{Alph}(u_{m+1}) = \mathrm{Alph}(\mathbf{s})$. Observe that, for $n < m$, the length of the longest return word to u_{n+1} is $|u_{n+1}| + 1$. Thus, successive occurrences of u_{n+1} in $\mathbf{s}$ are separated by at most one letter; in particular, $u_{n+1}xu_{n+1} \in F(\mathbf{s})$ for each $x \in \mathrm{Alph}(\mathbf{s}) \setminus \mathrm{Alph}(u_{n+1})$. On the other hand, for $n \geq m$, the length of the longest return word to u_{n+1} is $|r_{n+1}| = |u_{n+1}| - |u_p|$ where p is defined as above. Hence $|r_{n+1}| \leq |u_{n+1}|$, and therefore successive occurrences of u_{n+1} in $\mathbf{s}$ are either adjacent (i.e., $u_n u_n \in F(\mathbf{s})$) or they overlap. So, in light of Theorem 4.2 and the preceding remarks, we have essentially proved:

Theorem 4.5. *Any epistandard word* $\mathbf{s}$ *is quasiperiodic with smallest quasiperiod* $u_{m+1}u_p^{-1}$*, where* m *is minimal such that* $\mathrm{Alph}(x_1x_2\cdots x_m) = \mathrm{Alph}(\mathbf{s})$ *and* $p = \inf\{P(i) \mid i \geq m+1\}$*. Moreover, for all* $n \geq m$*,* u_{n+1} *is a quasiperiod of* $\mathbf{s}$*.* □

Remark 4.6. By equation (2.2), the smallest quasiperiod of $\mathbf{s}$ can be expressed as $u_{m+1}u_p^{-1} = h_{m-1}h_{m-2}\cdots h_{p-1}$. The length of this prefix is equal to that of the longest return word to u_{m+1}, which is given by $\mu_m(x)$ for some $x \in \mathrm{Alph}(x_{m+1}x_{m+2}\cdots)$.

Example 4.7. Consider the epistandard word directed by $(abc)^\omega$, namely the *Tribonacci word* (or *Rauzy word* [32]): $\mathbf{r} = abacabaabacababacabaabacabacabaabaca\cdots$.

Observe that $u_4 = abacaba$ is the shortest palindromic prefix of $\mathbf{r}$ such that $\mathrm{Alph}(u_4) = \{a, b, c\}$. The return words to u_4 in $\mathbf{r}$ are given by $\mu_3(x) = \psi_a\psi_b\psi_c(x)$ for each $x \in \{a, b, c\}$; explicitly: $\mu_3(a) = abacaba$, $\mu_3(b) = abacab$, $\mu_3(c) = abac$. So we see that successive occurrences of u_4 in $\mathbf{r}$ are either adjacent or overlap; hence $\mathbf{r}$ is $abacaba$-quasiperiodic. In fact, $u_4u_1^{-1} = \mu_3(a) = abacaba$ is the smallest quasiperiod of $\mathbf{r}$.

In general, the k-bonacci word, directed by $(a_1a_2\cdots a_k)^\omega$, is quasiperiodic with smallest quasiperiod u_{k+1}. This fact was also observed in [22] by noting that the k-bonacci word is generated by the morphism φ_k on $\{a_1, a_2, \ldots, a_k\}$ defined by $\varphi_k(a_i) = a_1a_{i+1}$ for all $i \neq k$, and $\varphi_k(a_k) = a_1$.

Even more, the following theorem completely describes *all* of the quaisperiods of an epistandard word.

Theorem 4.8. *Suppose* $\mathbf{s}$ *is an epistandard word directed by* $\Delta = x_1x_2x_3\cdots$ *and let* m *be minimal such that* $\mathrm{Alph}(x_1x_2\cdots x_m) = \mathrm{Alph}(\mathbf{s})$. *Then* $\mathbf{s}$ *is quasiperiodic and the set of all of its quasiperiods is given by* $\mathcal{Q} = \bigcup_{n\geq m}\mathcal{Q}_n$ *where* $\mathcal{Q}_n = \{u_{n+1}w^{-1} \mid w$ *is a suffix of* u_p *where* $p = \inf\{P(j) \mid j \geq n+1\}\}$. □

4.2.2 Some lemmas

We now state several lemmas concerning quasiperiodic episturmian words. These are used in the next section where we characterize the quasiperiodic episturmian words with respect to their directive words.

Lemma 4.9. *An episturmian word* $\mathbf{t}$ *is quasiperiodic if it is directed by a spinned word* $\breve{\Delta} = \breve{x}_1\breve{x}_2\breve{x}_3\cdots$ *with all spins ultimately equal to* 0. □

The following fact about return words is an important one to keep in mind.

Lemma 4.10. *Suppose* $\mathbf{s}$ *is an epistandard word and let* $\mathbf{t} \in \overline{\mathcal{O}}(\mathbf{s})$. *Then, for any factor* v *of* $\mathbf{s}$, $\mathcal{R}_v(\mathbf{s}) = \mathcal{R}_v(\mathbf{t})$. *That is, the return words to any factor* v *of an episturmian word* $\mathbf{t} \in \overline{\mathcal{O}}(\mathbf{s})$ *are the same as the return words to* v *as a factor of* $\mathbf{s}$. □

Lemma 4.11. *An episturmian word* $\mathbf{t}$ *is quasiperiodic if it is directed by* $\breve{\Delta} = v\breve{\mathbf{y}}$ *for some spinned infinite word* $\breve{\mathbf{y}}$ *and* $v \in \mathcal{A}^+$ *such that* $\mathrm{Alph}(v) = \mathrm{Alph}(\Delta)$. □

Definition 4.12. Suppose v is a recurrent factor of an infinite word $\mathbf{x}$ such that all of its return words have length at most $|v|$. Then any two successive occurrences of v in $\mathbf{x}$ are either adjacent or overlap each other, and we say that v is a ***totally overlapping*** factor of $\mathbf{x}$.

Note. If $\mathbf{x} \in \mathcal{A}^\omega$ is a quasiperiodic, then the quasiperiods of $\mathbf{x}$ are precisely its totally overlapping *prefixes*.

The next lemma, which can be viewed as an analogue of Theorem 4.8, gives the set of all totally overlapping factors of any episturmian word $\mathbf{t} \in \overline{\mathcal{O}}(\mathbf{s})$.

Lemma 4.13. *Suppose* $\mathbf{s}$ *is an epistandard word directed by* $\Delta = x_1x_2x_3\cdots$ *and let* m *be minimal such that* $\mathrm{Alph}(x_1x_2\cdots x_m) = \mathrm{Alph}(\mathbf{s})$. *Then the set of all totally overlapping factors of* $\mathbf{s}$ (*and hence any episturmian word* $\mathbf{t} \in \overline{\mathcal{O}}(\mathbf{s})$) *is given by* $\mathbb{O} = \bigcup_{n\geq m}\mathbb{O}_n$ *where*

$$\mathbb{O}_n = \{v \in F(u_{n+1}) \setminus F(u_n) \mid |v| \geq |u_{n+1}| - |u_p| \text{ where } p = \inf\{P(j) \mid j \geq n+1\}\}. \qquad \square$$

Lemma 4.13 yields the following trivial characterization of the quasiperiodic episturmian words.

Corollary 4.14. *An episturmian word* $\mathbf{t} \in \overline{\mathcal{O}}(\mathbf{s})$ *is quasiperiodic iff some* $v \in \mathbb{O}$ *is a prefix of* $\mathbf{t}$. □

4.3 Non-quasiperiodic episturmian words

As a consequence of Theorems 4.5-4.8, the problem of determining which episturmian words are not quasiperiodic now reduces to considering only the non-epistandard ones. Moreover, as any periodic infinite word is quasiperiodic [26, Proposition 1], we may also let aside the periodic episturmian words.

4.3.1 Regular wavy words

Following the terminology in [20], a spinned version $\breve{\Delta}$ of Δ is said to be *wavy* if $\breve{\Delta}$ contains infinitely many spins 0 and 1. The *opposite* $\hat{w}$ of a finite or infinite spinned word $\breve{w}$ is obtained from $\breve{w}$ be exchanging all spins in $\breve{w}$. If $v \in \mathcal{A}^+$, then its opposite $\bar{v} \in \bar{\mathcal{A}}^+$ has all spins 1.

Remark 4.15. By Proposition 3.11 in [18], if a spinned version $\breve{\Delta}$ of Δ has infinitely many spins 0, then $\breve{\Delta}$ directs *exactly one* episturmian word. Accordingly, if $\breve{\Delta}$ is wavy or has all spins ultimately equal to 0, then there exists a unique episturmian word $\mathbf{t}$ directed by $\breve{\Delta}$. Moreover, $\mathbf{t}$ begins with the left-most letter in Δ having spin 0 in $\breve{\Delta}$ (by properties of episturmian morphisms). On the other hand, if all spins of $\breve{\Delta}$ are ultimately equal to 1, then there are exactly $|\mathrm{Ult}(\Delta)|$ episturmian words directed by $\breve{\Delta}$ and their first letters are those in $\mathrm{Ult}(\Delta)$.

We prove the following useful fact about directive words.

Lemma 4.16. *Any episturmian word has a wavy directive word.* □

From Section 3.2, recall that the (unique) aperiodic episturmian word directed by $\bar{\Delta}_x$ (with $x \in \mathrm{Ult}(\Delta)$ and $|\mathrm{Ult}(\Delta)| \geq 2$) is an infinite Lyndon word for any order with $x = \min(\mathcal{A})$. As such, since any Lyndon word is not quasiperiodic [23, Corollary 6.3], we see that $\bar{\Delta}_x$ directs a non-quasiperiodic episturmian word. Such a wavy directive word is 'regular' in the sense of the following definition.

Definition 4.17. A spinned version $\breve{w}$ of a finite or infinite word w is said to be ***regular*** if, for each letter $x \in \mathrm{Alph}(w)$, all occurrences of $\breve{x}$ in $\breve{w}$ have the same spin (0 or 1).

For example, $a\bar{b}aa\bar{c}\bar{b}$ and $(a\bar{b}c)^\omega$ are regular, whereas $a\bar{b}a\bar{a}\bar{c}b$ and $(a\bar{b}\bar{a})^\omega$ are not regular. More generally, Δ, $\bar{\Delta}_x$ and their opposites are regular.

Example 4.18. If $\mathrm{Alph}(\Delta) = \mathrm{Ult}(\Delta) = \{a, b\}$, then the regular 'spinned' versions of Δ are itself, its opposite $\bar{\Delta}$, and the wavy words $\bar{\Delta}_a$ and $\bar{\Delta}_b$. The latter two regular wavy words direct the Sturmian Lyndon words $a\mathbf{s}$ and $b\mathbf{s}$ that are equivalent to the standard word $\mathbf{s}$ directed by Δ.

Remark 4.19. A *regular wavy* word $\breve{\Delta}$ necessarily has $|\mathrm{Ult}(\Delta)| \geq 2$; that is, there exist letters $a, b \in \mathrm{Ult}(\Delta)$, $a \neq b$, such that all spins of a (resp. b) in $\breve{\Delta}$ are 0 (resp. 1). Thus, any regular wavy word directs an *aperiodic* episturmian word.

The relevance of regular wavy words is highlighted by the following lemma and theorem.

Lemma 4.20. *If $\breve{\Delta}$ is a regular wavy word, then $\breve{\Delta}$ is the* unique *directive word for exactly one (aperiodic) episturmian word.* □

Theorem 4.21. *If an episturmian word $\mathbf{t}$ is directed by a regular wavy word, then $\mathbf{t}$ is not quasiperiodic.* □

Example 4.22. With $\Delta = (abcd)^\omega$, the regular wavy words $\bar{\Delta}_a$, $\bar{\Delta}_b$, $\bar{\Delta}_c$, $\bar{\Delta}_d$, $(\bar{a}\bar{b}cd)^\omega$, $(\bar{a}b\bar{c}d)^\omega$, $(\bar{a}bc\bar{d})^\omega$ and their opposites direct non-quasiperiodic episturmian words that are equivalent to the 4-bonacci word.

Theorem 4.21 shows that there is a much wider class of episturmian words that are not quasiperiodic, besides those that are infinite Lyndon words.

The following example shows that an 'ultimately regular' wavy word does not necessarily direct a non-quasiperiodic episturmian word.

Example 4.23. Recall the Tribonacci word $\mathbf{r}$ from Example 4.7. Observe that the ultimately regular wavy word $\breve{\Delta} = aadbc\bar{d}(a\bar{b}\bar{c})^\omega$ directs the quasiperiodic episturmian word $\mu_{aadbc\bar{d}}(a\mathbf{r})$, which has smallest quasiperiod $u_6u_2^{-1} = aadaabaadaacaadaabaada$ (the same as the epistandard word directed by $aadbcd(abc)^\omega$). This example illustrates Lemma 4.11. However, if we remove d from the prefix of $\breve{\Delta}$, then the ultimately regular wavy word $aabc\bar{d}(a\bar{b}\bar{c})^\omega$ directs the non-quasiperiodic episturmian word $\mu_{aabc\bar{d}}(a\mathbf{r})$. Notice that the prefix $aabc$ (with all spins 0) of the latter directive word does not contain all of the letters in its alphabet.

4.3.2 Characterizations

Theorem 4.24. *An episturmian word is quasiperiodic if and only if it has a directive word of the form $\breve{w}v\breve{\mathbf{y}}$, for some spinned infinite word $\breve{\mathbf{y}}$ and words w, v with* $\mathrm{Alph}(v) = \mathrm{Alph}(v\mathbf{y})$. □

Finally, we state the announced characterization of the non-quasiperiodic episturmian words, which is simply a reformulation of the above theorem.

Theorem 4.25. *An episturmian word is not quasiperiodic if and only if it does not have a directive word of the form $\breve{w}v\breve{\mathbf{y}}$, where $\breve{\mathbf{y}}$ is a spinned infinite word and w, v are words with* $\mathrm{Alph}(v) = \mathrm{Alph}(v\mathbf{y})$. □

Example 4.26. Let us demonstrate Theorems 4.24–4.25 using the ever-so popular Tribonacci word $\mathbf{r}$. By Theorem 3.6, we know that $a\mathbf{r}$ is an infinite Lyndon word, and hence is not quasiperiodic; nor is the 'ultimately Lyndon' episturmian word $\psi_c(a\mathbf{r})$, directed by $c\bar{\Delta}_a = c(a\bar{b}\bar{c})^\omega$. However, $\mu_{bc}(a\mathbf{r})$, $\mu_{cb}(a\mathbf{r})$, $\mu_{\bar{a}\bar{d}}(a\mathbf{r})$ are *quasiperiodic* episturmian words, respectively directed by

$$bc(a\bar{b}\bar{c})^\omega = bca\bar{b}(\bar{c}a\bar{b})^\omega, \quad cb(a\bar{b}\bar{c})^\omega = cba\bar{b}(\bar{c}a\bar{b})^\omega, \quad \bar{a}\bar{d}(a\bar{b}\bar{c})^\omega.$$

Notice that the first two spinned words take the form $v\breve{\mathbf{y}}$ where $\breve{\mathbf{y}}$ is regular wavy and $\mathrm{Alph}(v) = \mathrm{Alph}(v\mathbf{y})$. The last one $\breve{\Delta} = \bar{a}\bar{d}(a\bar{b}\bar{c})^\omega$ directs the same episturmian word as $\hat{\Delta} = adabc\bar{a}(\bar{b}\bar{c}a)^\omega$ (since $\breve{\Delta}$ and $\hat{\Delta}$ are *block-equivalent* [20]), and $\hat{\Delta}$ takes the form given by Theorem 4.24.

5 Self-episturmian numbers

Let $\beta > 1$ be a real number. We denote by $d_\beta(x)$ the *greedy β-expansion* (or *Rényi expansion in base β*) [33] of a real number $x \in [0,1]$, which is an infinite sequence $(v_i)_{i\geq 1}$ of non-negative integers on the alphabet $\mathcal{A}_\beta := \{0, 1, \ldots, \lfloor\beta\rfloor\}$ (possibly degenerating to infinitely many 0's).

We say that a real number $\beta > 1$ is *self-episturmian* if $d_\beta(1)$ is an $\mathcal{A}_\beta$-strict episturmian sequence. In this last section, we prove a simple characterization of self-episturmian numbers, which is an easy consequence of some of our results in Section 3. In [15], we show further that self-episturmian numbers are transcendental.

5.1 Self-episturmianicity

Every real number $\beta > 1$ is characterized by its Rényi expansion of 1; however, not every sequence $(v_i)_{i\geq 1}$ of non-negative integers is equal to $d_\beta(1)$ for some β. In [27], Parry gave the following necessary and sufficient condition for a sequence to be a greedy β-expansion of 1 for some β.

Proposition 5.1. [27] *A sequence $\mathbf{v} = (v_i)_{i\geq 0} \in \mathcal{A}_\beta^\omega$ is a greedy β-expansion of 1 for some β if and only if $\mathrm{T}^k(\mathbf{v}) < \mathbf{v}$ for all $k \geq 1$, in which case β is unique.* □

Remark 5.2. In other words, a sequence $\mathbf{v} = (v_i)_{i\geq 0} \in \mathcal{A}_\beta^\omega$ is equal to $d_\beta(1)$ for some β if and only if $\mathbf{v}$ is lexicographically greater than all of its proper suffixes (and hence dominates its shift space), i.e., if and only if $\mathbf{v}$ is an *infinite anti-Lyndon word* for the lexicographic order induced by $0 < 1 < \cdots < \lfloor\beta\rfloor$.

Theorem 5.3. *A real number $\beta > 1$ is self-episturmian if and only if $d_\beta(1) = \lfloor\beta\rfloor\mathbf{s}$ where $\mathbf{s}$ is an $\mathcal{A}_\beta$-strict epistandard sequence.* □

The above theorem generalizes one of the main results in [8], where *self-Sturmian numbers* were first introduced and studied.

5.2 Univoque self-episturmian numbers: a conjecture

A *univoque number* is a real number $\beta > 1$ such that 1 has a unique expansion in base β. See for instance [2] and references therein, or more recently [3] in which univoque *Pisot numbers* in $(1,2)$ are studied.

Allouche [1] has characterized the univoque self-Sturmian numbers in $(1,2)$. More generally, we make the following conjecture for univoque self-episturmian numbers.

Conjecture 5.4. *A real number $\beta > 1$ is univoque and self-episturmian if and only if $d_\beta(1) = \lfloor\beta\rfloor\mathbf{s}$ where $\mathbf{s}$ is an $\mathcal{A}_\beta$-strict epistandard sequence beginning with $\lfloor\beta\rfloor$.*

Acknowledgements

Many thanks to Jacques Justin for his helpful comments on a preliminary version of this paper. Thanks also to Gwénaël Richomme for sending me a copy of his paper [35].

References

[1] J.-P. Allouche, A note on univoque self-Sturmian numbers, Preprint, arXiv:math/0612816v1.

[2] J.-P. Allouche, M. Cosnard, Non-integer bases, iteration of continuous real maps, and an arithmetic self-similar set, *Acta Math. Hungar.* **91** (2001) 325–332.

[3] J.-P. Allouche, C. Frougny, K.G. Hare, On univoque Pisot numbers, *Math. Comp.* (to appear).

[4] J.-P. Allouche, A. Glen, Extremal properties of (epi)sturmian sequences and distribution modulo 1 (in preparation).

[5] A. Apostolico, M. Crochemore, String pattern matching for a deluge survival kit, in *Handbook of Massive Data Sets, Massive Comput.*, vol. 4, edited by J. Abello, P.M. Pardalos, M.G.C. Resende, Kluwer Acad. Publ., Dordrecht, 2002.

[6] A. Apostolico, A. Ehrenfeucht, Efficient detection of quasiperiodicities in strings, *Theoret. Comput. Sci.* **119** (1993) 247–265.

[7] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, *Bull. Soc. Math. France* **119** (1991) 199–215.

[8] D.P. Chi, D.Y. Kwon, Sturmian words, β-shifts, and transcendence, *Theoret. Comput. Sci.* **321** (2004) 395–404.

[9] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* **7** (1973) 138–153.

[10] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* **183** (1997) 45–82.

[11] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001) 539–553.

[12] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* **179** (1998) 89–101.

[13] A. Glen, Powers in a class of $\mathcal{A}$-strict standard episturmian words, *Theoret. Comput. Sci.* **380** (2007) 330–354.

[14] A. Glen, A characterization of fine words over a finite alphabet, *Theoret. Comput. Sci.* (to appear).

[15] A. Glen, A note on self-episturmian numbers (in preparation).

[16] A. Glen, J. Justin, G. Pirillo, Characterizations of finite and infinite episturmian words via lexicographic orderings, *European J. Combin.* (in press), doi:10.1016/j.ejc.2007.01.002.

[17] C. Holton, L.Q. Zamboni, Descendants of primitive substitutions, *Theory Comput. Syst.* **32** (1999) 133–157.

[18] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* **276** (2002) 281–313.

[19] J. Justin, G. Pirillo, On a characteristic property of Arnoux-Rauzy sequences, *Theoret. Inform. Appl.* **36** (2002) 385–388.

[20] J. Justin, G. Pirillo, Episturmian words: shifts, morphisms and numeration systems, *Internat. J. Found. Comput. Sci.* **15** (2004) 329–348.

[21] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words, *Theoret. Inform. Appl.* **34** (2000) 343–356.

[22] F. Levé, G. Richomme, Quasiperiodic infinite words: some answers, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **84** (2004) 128–138.

[23] F. Levé, G. Richomme, Quasiperiodic Sturmian words and morphisms, *Theoret. Comput. Sci.* **372** (2007) 15–25.

[24] M. Lothaire, *Combinatorics On Words, Encyclopedia of Mathematics and its Applications*, vol. 17, *Addison-Wesley*, Reading, Massachusetts, 1983.

[25] M. Lothaire, *Algebraic Combinatorics On Words, Encyclopedia of Mathematics and its Applications*, vol. 90, *Cambridge University Press*, U.K., 2002.

[26] S. Marcus, Quasiperiodic infinite words, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **82** (2004) 170–174.

[27] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960) 401–416.

[28] G. Pirillo, Inequalities characterizing standard Sturmian words, *Pure Math. Appl.* **14** (2003) 141–144.

[29] G. Pirillo, Inequalities characterizing standard Sturmian and episturmian words, *Theoret. Comput. Sci.* **341** (2005) 276–292.

[30] G. Pirillo, Morse and Hedlund's skew Sturmian words revisited, *Ann. Comb.* (to appear).

[31] N. Pytheas Fogg, *Substitutions In Dynamics, Arithmetics And Combinatorics, Lecture Notes in Mathematics*, vol. 1794, *Springer-Verlag*, Berlin, 2002.

[32] G. Rauzy, Nombres algébriques et substitutions, *Bull. Soc. Math. France* **110** (1982) 147–178.

[33] A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hungar.* **8** (1957) 477–493.

[34] G. Richomme, Conjugacy and episturmian morphisms, *Theoret. Comput. Sci.* **302** (2003) 1–34.

[35] G. Richomme, Lyndon morphisms, *Bull. Belg. Math. Soc. Simon Stevin* **10** (2003) 761–785.

[36] G. Richomme, On morphisms preserving infinite Lyndon words, *Discrete Math. Theor. Comput. Sci.* **9** (2007), 89–108.

[37] R.N. Risley, L.Q. Zamboni, A generalization of Sturmian sequences: combinatorial structure and transcendence, *Acta Arith.* **95** (2000) 167–184.

[38] P. Veerman, Symbolic dynamics and rotation numbers, *Physica A*, **134** (1986) 543–576.

[39] P. Veerman, Symbolic dynamics of order-preserving orbits, *Physica D*, **29** (1987) 191–201.